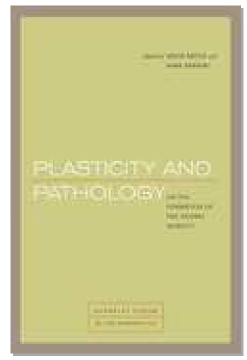
Plasticity and Pathology

Bates, David , Bassiri, Nima



Published by Fordham University Press

Bates, David . and Bassiri, Nima .
Plasticity and Pathology: On the Formation of the Neural Subject.
New York: Fordham University Press, 2015.
Project MUSE. Web. 7 Dec. 2015. http://muse.jhu.edu/.

David Bates

7 Automaticity, Plasticity, and the Deviant Origins of Artificial Intelligence

THE CONTEMPORARY BRAIN is largely a digital brain.[1] Not only do we study the brain through virtual technologies that rely on digital visualizations, but the brain's very activity is often modeled by a digital simulation.[2] And the brain is, in many different ways, still understood to be a digital machine; it is a kind of neural computer.[3] The legacy of artificial intelligence (AI) still persists in contemporary neuroscience and cognitive science.[4] The two competing projects to "map" the brain in Europe and the United States clearly reveal the discursive and conceptual connections between computers and neurophysiology—and *neuropathology*. As the European Union Human Brain Project website puts it, understanding the human brain will allow us to "gain profound insight into what makes us human, develop new treatments for brain disease and build revolutionary new computing technologies."[5] Similarly the Connectome project initiated by the National Institutes of Health and supported by President Obama with substantial funding, reveals the essential link between new research technologies, the technologized vision of the brain itself, and

pathologies: “Long desired by researchers seeking new ways to treat, cure, and even prevent brain disorders, this picture will fill major gaps in our current knowledge and provide unprecedented opportunities for exploring exactly how the brain enables the human body to record, process, utilize, store, and retrieve vast quantities of information, all at the speed of thought.”[6]

It would seem that the recent intensification of interest in the inherent plasticity of the brain—its developmental openness, its always evolving structure in the adult phase, and its often startling ability to reorganize itself after significant trauma—puts considerable pressure on the technological conceptualizations of the brain that assume a complex but definite automaticity of operation. Indeed the concept of plasticity has been heralded as a counter to the machinic understanding of the brain, most notably by the philosopher Catherine Malabou.[7] However, it is now the case that the neurophysiological phenomenon of brain plasticity is rapidly becoming assimilated to computational models and digital representations of the nervous system. In the field of computational developmental neuroscience, for example, the brain is figured as a *learning machine* that automatically constructs, according to set algorithms, connective webs that are dependent on the specific “experience” of the network. These models are all ultimately derived from the seminal theory of the neurophysiologist Donald Hebb. As he famously explained in his 1949 book *The Organization of Behavior*, synaptic connections between neurons are strengthened with use; the theory is often reduced to the paraphrase “Neurons that fire together wire together.” The processes governing the determination of the plastic brain as it experiences the world are obviously much more complex, but the basic principle still holds. Therefore even the contingent determination of the plastic brain can, it is thought, be rigorously modeled by a virtual computer simulation.

This doubling of the brain by its digital other has in turn affected the technological domain of computing itself. Attempts

to model the cortex of animal brains with "synaptic" architectures, for example, are framed as experimental investigations of the neural organization itself; it is said that this is "a significant step towards unravelling the mystery of what the brain computes," which in turn, the researchers claim, open "the path to low-power, compact neuromorphic and synaptronic computers of tomorrow."[8] Digital visions of the plastic brain have stimulated the invention of new computational architectures. IBM, for example, has designed a chip that mimics the networked connectivity of neural systems. The chip has "digital neurons" that can rewire synapses based on their inputs.[9] A rival Intel project also promises programmable "synapses" and the imitation of "integrating-neurons."[10] These neurosynaptic chips are part of a broader investigation of what is being called "cognitive computing," whose aim "is to develop a coherent, unified, universal mechanism inspired by the mind's capabilities." That is, the researchers "seek nothing less than to discover, demonstrate, and deliver the core algorithms of the brain and gain a deep scientific understanding of how the mind perceives, thinks, and acts." Here the brain has become an algorithmic learning *machine*; it is a mirror of the very technology that represents it. Analysis of the brain's networked organization will, it is claimed, "lead to novel cognitive systems, computing architectures, programming paradigms, practical applications, and intelligent business machines."[11]

This mirroring has in fact a long history—arguably as long as the history of computing itself. The goal of early AI research was twofold: to produce an intelligent simulation and, by doing so, test hypotheses about the functions of the human mind. Underlying this project was the assumption that the mind or brain would be amenable to such analysis. As researchers wrote in 1954, with respect to their own effort at cortical modeling, "No matter how complex the organization of a system . . . it can always be simulated as closely as desired by a digital computer as long as its rules of organization are known."[12] This was, they note, the implication

of Alan Turing's early conceptualization of the universal digital computer as a machine that can (virtually) imitate any other discrete-state machine.

The automaticity of the brain's operation (even in its most radically plastic guise) is to a great extent a consequence of the historical codevelopment of computer technologies, AI, and the cognitive sciences. The question of *autonomy* in this framework of automaticity is drained of all potential meaning. And so, once plasticity has been fully integrated into the computational and neurophysiological models of the brain, resistance to the total automatization of human thinking cannot simply rely on romanticized concepts of selfhood or philosophical critiques of materialism; we must focus instead on the historical and conceptual foundations of the digital brain itself. Resistance can be generated, that is, through a critical history of *automaticity*. By returning to the threshold of the digital age, that moment when the modern computer was first being conceptualized (and the ideas and practices of AI simultaneously set in motion), we can see that the digital was not, at the beginning, fully aligned with automaticity. Indeed, although it has not received much attention, key developers of early computer technologies were explicitly trying to imitate, with their new thinking technologies, a more radical openness, a more unpredictable plasticity within the nervous system—a subject that was, we will see, very much alive in early twentieth-century neurophysiology and neurologically informed experimental psychology. At the same time that some cyberneticians were claiming that the brain was just an automatic calculator like the computer, crucial figures in the history of computing and cybernetics immediately recognized the importance of the plasticity of the brain for the project of AI: the plastic brain, it was thought, offered the possibility of modeling creative, unpredictable leaps of human intelligence, capacities that went *beyond* the relentlessly automatic performance of rigid functional mechanisms or habitual behaviors. It is therefore significant that the

neurophysiological discourses of plasticity in the period were intimately linked to the disorders and crises of the diseased or injured brain. Constructing a plastic computational machine at the dawn of the digital age therefore entailed, I will argue, the invention of a *machinic pathology*.

Recuperating this historical moment will offer a new perspective on our contemporary "digital brain." We need not be reduced to mere learning machines, largely unconscious of our own cognitive processes, where any experience of freedom and contingency of thinking can be exposed as some kind of Nietzschean illusion. The human brain was understood to be a special kind of genuinely open system that determined itself yet was also capable of defying its own automaticity in acts of genuine creativity. The originators of the digital computer were explicitly inspired by this neurophysiological concept of plasticity in their efforts to model the abilities of truly intelligent beings.

HISTORIES OF AI usually trace the lineage back through figures who attempted to simulate thinking in some kind of automatic machinery. It could be said that René Descartes's philosophical and physiological writings gave us a vision of the first modern automaton—that is, a *thinking* machine. While he of course resisted the ultimate implications of his own systematic mechanization of the human and animal body, Descartes pointed the way to a mechanical understanding of cognition when he gave a comprehensive description of the nervous system and the essential role it played not just in governing all animal behavior but also, more importantly, in producing the vast majority of *human* thinking and action, the routine cognition of the everyday.

Yet a closer look at Descartes's writing reveals a theory that is somewhat more complex than a merely mechanical vision of the body's operations. Descartes showed how the nerves and the brain were an information system, remarkable for its flexibility and adaptability. While the nervous system was a material struc-

ture, it was nonetheless plastic and modifiable. As a space for integrating information, the brain was, he wrote, "soft and pliant" ("molle et pliante") and capable therefore of being imprinted with memories and of acquiring reflexes.[13] The implication was that the cultural determination of the individual—through language, culture, history—took place within the soft architecture of the brain itself. As he remarked in the *Discourse on Method*, "I thought too, how the same man, with the same mind, if brought up from infancy among the French or Germans, develops otherwise than he would if he had always lived among the Chinese or cannibals."[14] The open brain was determined by the physical flows that were produced by the organs of sense and transmitted through the conduits of the nerves. Descartes's automaton was no mere clockwork mechanism but an open site of perpetual organization and reorganization occasioned by information received through the sensory systems.[15]

It is true that the long history of automated thinking technologies shows that the idea of a "soft and pliant" system was overshadowed by the hard mechanisms of the industrial age. From the semi-automatic calculators of Pascal and Leibniz to Charles Babbage's Difference Engine (the first example of a truly automatic form of artificial thought) and the unfinished general computer he called the Analytic Engine, and then on to the "logic pianos" of the late nineteenth century (built by, for example, William Jevons and Allan Marquand), models of human reasoning were concretely instantiated by machines with precise and predictable operations. It is hardly surprising, then, that these early examples of artificially mechanized reasoning would greatly influence the conceptualization of human cognition. For example, Charles Sanders Peirce, the American pragmatist philosopher, reflecting on logic machines, noted that the performance of a deductive rational inference had "no relation to the circumstance that the machine happens to work by geared wheels, while a man happens to work by an ill-understood arrangement of brain-cells."[16]

Peirce himself had the insight (many years before Claude Shannon) that these logical relations could even be performed by means of electrical circuits.[17] With this step we are on the very edge of modern computing and, with it, the effort to understand the brain itself as a machine for thinking, constructed from so many neural "switches."

We should pause, however, to remark that in the period before digital computing, knowledge of the brain was not at all congruent with the mechanistic paradigm of automatism. By the late nineteenth century the idea that the brain was made up of multiple, localizable components was giving way to models of the brain that emphasized its distributed structural organization. The great British neurologist John Hughlings Jackson, for example, used extensive clinical investigation of neural disorders to argue that the brain had many levels of organization due to the gradual development of the organ in the course of evolution. What is interesting is that Hughlings Jackson wanted to demonstrate that the highest intellectual capacities were associated with the *least* organized, least "evolved," and what he called "least automatic" cortical structures. The most automatic functions governed our basic physiological systems, while the relatively open and undetermined cortex was associated with complex thinking and discursive language.

William James, perhaps the most influential figure in the synthesis of psychology and neuroscience at the turn of the century, was very much interested in how the human mind was driven by habitual and unconscious cognitions. These were not given but rather acquired. The brain was at once a site of openness and a space of artificial mechanisms: "The phenomena of habit in living beings are due to the plasticity of the organic materials of which their bodies are composed." "Plasticity," wrote James, "means the possession of a structure weak enough to yield to an influence, but strong enough not to yield all at once." For James, as for Peirce, the organismic forms of life were essentially plastic since

they were the product of evolutionary change, and the nervous system was the privileged site for this formation and reformation as the organism adapted to its changing environmental conditions. "Organic matter, especially nervous tissue, seems endowed with a very extraordinary degree of plasticity," he noted, and this explained the way beings could become automatons of a sort—but never true automata, precisely because plasticity never entirely disappeared from the brain and nervous system.[18]

Drawing on the evolutionary theories of figures such as Herbert Spencer and Hughlings Jackson, James took up the idea that the human capacity for intuition and insight was most likely linked not to some specific organ of intelligence but in fact to the instability and indetermination of the higher centers of the brain. James suggested that "the very vagueness [of the cerebral hemispheres] constitutes their advantage. . . . An organ swayed by slight impressions is an organ whose natural state is one of unstable equilibrium." This "plasticity" of the higher brain made it unpredictable; it was "as likely to do the crazy as the sane thing at any given moment." But James went even further, noting that an injured brain, missing component parts that have been damaged or destroyed entirely, might be seen as "virtually a new machine." This new machine may initially perform abnormally, with psychopathological symptoms, but as neurological cases repeatedly demonstrated, this new brain often found its way back to normality. "A brain with part of it scooped out is virtually a new machine, and during the first days after the operation functions in a thoroughly abnormal manner. As a matter of fact, however, its performances become from day to day more normal, until at last a practiced eye may be needed to suspect anything wrong."[19] James's colleague Peirce had similarly linked pathology with novelty. When our brain is injured, Peirce noted, we act in "new ways." Usually this entails a state of mental illness, but Peirce was willing to admit that in special cases "disease of the brain may cause an improvement in the general intelligence."[20] The inherent

plasticity of the brain revealed itself best in the injured brain, but the ability to reorganize and defend against damage was linked to the cognitive flexibility of human minds. Intelligence was, in a sense, considered to be a consequence of a certain disorganization and unpredictability, and potentially even pathological disorder might explain the leaps of a genius intelligence.

Clinical and experimental research on the brain in the early twentieth century was systematically exploring the ability of the brain to *reorganize* in the face of challenges—including the radical challenge of grave injury. The shock of disorder opened up the possibility of a new form of order that was not explicable in merely mechanical terms. As Constantin von Monakow put it in a book on brain localizations, the disruption of part of the brain led to a more general "shock" of the system (what he called *diaschisis*). "Any injury suffered by the brain substance will lead (just as lesions in any other organ) to a struggle [*Kampf*] for the preservation of the disrupted nervous function, and the central nervous system is always (though not always to the same degree) prepared for such a struggle."[21] The pathological turn awakened a total response, aimed not at a simple return to the original order but rather to an order that reestablished stable functioning in a new form. As von Monakow wrote (in a book cowritten with the French neurologist R. Mourgue), "It is a question of combat, of an active struggle for the creation of a new state of affairs, permitting a new adaption of the individual and its milieu. The whole organism intervenes in this struggle."[22]

Interwar neurological theory produced many theoretical models that emphasized the plastic nervous system that could adjust itself to constantly changing circumstances, and even to radical damage of the brain.[23] One influential figure in this field, the American psychologist Karl Lashley, sought to disprove localization theories of memory by systematically removing pieces of the brain of animal subjects and demonstrating the persistence of learned behaviors in mazes and other environments. Surprisingly the test animals were

often still able to run the mazes efficiently, implying that the brain must have a way to restructure itself to compensate for the missing tissue. Lashley saw this capacity as continuous with the adaptive unity of the brain as a complex system, a system that integrated activity throughout its component parts: "The whole implication of the data is that the 'higher level' integrations are not dependent upon localized structural differentiations but are a function of some general, dynamic organization of the entire cerebral system." Lashley's term for this flexibility was *equipotentiality*, alluding to his theory that because the brain was a dynamic and ever-changing entity, it could *reorganize* itself when challenged with new obstacles or internal failures.[24] As he later wrote, "I have been struck by the fact that even very extensive destruction of brain tissue does not produce a disorganization. Behavior becomes simplified but remains adaptive." For Lashley intuitive, insightful, intellectual activity was dependent on this feature of the brain and nervous system—"its plastic and adaptive" nature.[25]

Perhaps the most important figure in this history of plasticity was Kurt Goldstein, who synthesized his own extensive clinical research on neurological defects in brain-injured patients with broader streams of thinking about organismic life itself. Using neurological concepts such as *diaschisis*, taken from von Monakow's work, as well as German holistic philosophy, especially Gestalt theory, Goldstein defined the organism less as a fixed organization or structure and more as a dynamic configuration constantly struggling to maintain coherent unity when challenged by the ever-changing conditions of life. Pathological states of being exhibited their own peculiar characteristics in neurological patients. When an organism's responses to the environment were "constant, correct, adequate," it could survive in its milieu. But in a condition of shock the organism was often led to "disordered, inconstant, inconsistent" actions, which created a "disturbing aftereffect."[26]

However, as Goldstein would argue, organisms have the capacity to modify themselves in the face of this disordered behavior,

which he called a "catastrophe reaction." The organism's ability to reorganize in response to shock was, according to Goldstein, just one way that this organism sought unity in moments of disruption. The catastrophes that were labeled pathological were in fact continuous with the dynamic of a healthy, normal body as it struggled to maintain its equilibrium. That is, the organism was always seeking unity and order: "Normal as well as abnormal reactions ('symptoms') are only expressions of the organism's attempt to deal with certain demands of the environment."[27]

And so Goldstein writes that even the normal, healthy life of the organism can be considered a series of what he calls "slight catastrophes" (*leichter Katastrophenreaktionen*), where inadequacies are first confronted and then "new adjustments" or "new adequate milieu" are sought to respond to this lack. The more serious catastrophic breakdown differs only in the scale and intensity of the reaction. The whole organism is really a potential unity always falling into states of shock, and it must, over and over again, creatively establish new order to overcome these shocks: "If the organism is 'to be,' it always has to pass again from moments of catastrophe to states of ordered behavior."[28] The essential plasticity of the organism was of course most clearly visible in the brain because it was the center of organization and integration and therefore the site of reorganization and reconstitution.

However, the same structures of crisis and reorder can be tracked in interwar psychology in theories of creative and adaptive thinking. In Gestalt psychology most clearly, genuine intelligence was defined as the ability to solve difficult problems with new perspectives that reorganized the experience of the subject to produce new arrangements. The word used for this sudden realization was *insight*. In his famous ape studies conducted during World War I, Wolfgang Köhler aimed to uncover the primordial conditions of intelligence, using primates as a way into the essence of human intelligence before it was overlaid with language and cultural knowledge. Köhler showed how insight (*Einsicht*) was

achieved when the mind freed itself from one interpretation of the situation to discover a new one that resolved the tension—in these cases, to see objects as tools that could lead to food that had been placed out of direct reach. For Köhler intelligence was revealed by the ability to "detour" away from the direct path in order to understand how to circumvent obstacles.[29] Psychological insight had its "isomorphic" counterpart in the organism itself: "Human beings construct machines and compel natural forces to operate in defined and controllable ways. A system of this sort is one of 'fixed' connections in the sense of analytical mechanics." But a *dynamic* system was capable of reorganization and adaptation. Köhler's example was the developing organism: "Many embryos can be temporarily deformed (artificially) in ways utterly incongruous with their racial history and nevertheless regain by a wholly other route the same developmental stage reached by another, untouched embryo."[30]

The investigation into "productive thinking" emphasized the novelty of analogical extensions of established knowledge into radically new zones of understanding.[31] Köhler's Gestalt colleague Kurt Koffka, for example, argued that the mind was not limited to combining and recombining the "reaction pathways" and memories that it already possessed. The mind had, he said, a fundamental "plasticity" that allowed for radical novelty and therefore real progress in thinking.[32] The Swiss experimental psychologist Edouard Claparède perfectly captured this relationship between automatisms of habitual thinking and radical liberty characteristic of intelligence; he also echoed Goldstein's idea of the catastrophic reaction: "The rupture of equilibrium, when reflexes or habits are not available to intervene, is not reestablished automatically, and we are momentarily disadapted. It is intelligence which takes on the task of readapting us."[33]

HERE, AT THE threshold of the computer age, it was understood that the brain's open structure precluded any easy reduction of human cognition to some forms of automatic "machinery." Yet the most

advanced computing devices of the period were analog machines, such as Vannevar Bush's Differential Analyzer (ca. 1930) and the hybrid electromechanical Mark I at Harvard (1940–44). However complex these technological machines were, they clearly incarnated the principle of *automaticity*: constructed as a series of physically coupled moving elements, these analog devices were arranged to represent directly the relationships of various mathematical or other determined functions. This seemed to preclude any serious form of AI that could mimic intelligent life. As Vladimir Jankélévitch put it in 1931 in a book on Bergson, "With a perfect machine there is never any deception, but also never any surprise. None of those miracles which are, in a way, the signature of life."[34]

So when Turing first imagined the "universal" digital computer in a 1936 mathematical paper it was a strange interloper in the field of automatic computing precisely because it had *no intrinsic organization* of its own. As a simple machine with nothing more than the capacity to manipulate two symbols on a moving tape, this *digital* computer was defined as a radically open machine whose sole purpose was to take on the configurations of other discrete-state machines. What Turing invented was a machine that automatically mimicked (virtually, in coded representations) the successive states of *another* automatic machine. Soon enough, of course, the Turing machine became much more than that. The binary logic of the digital computer was quickly applied to synaptic connectivity in the brain. Turing's main point—that a digital computer's operation was governed by its *logical* and not physical organization—only strengthened the analogy of computer and brain in this period. For some the brain was in essence a digital computer instantiated by the sequential firing of neurons, which were analogous to the mechanical switching of relays or the processing of electrical impulses. The physical substrate of the Turing machine was irrelevant to its logical operation.[35]

Critics of the mechanistic worldview zeroed in on the rigid automaticity implied by the computer analogy. Machines, wrote

Georges Canguilhem in a 1948 essay, could only affirm rational "norms" of consistency and predictability. In contrast living beings were capable of *improvisation*; in its creative drive to survive, "life tolerates monstrosities." Pointing to the great plasticity of the nervous system, Canguilhem noted that if a child suffers a stroke that destroys an entire half of the brain, that child would not suffer aphasia (as is the case with many brain injuries later in life) because the brain reroutes language performance to other regions in order to compensate for the damage. Organisms are stable as *unities* precisely because their organization is *not* fixed into any one rigid structure; they are open, and thus equipped to surmount even a traumatic loss of functions in some cases. However, as Canguilhem declared, "There is no machine pathology."[36] And of course for Canguilhem the pathological state of illness revealed the living being's power to create new norms in crisis situations.[37] Michael Polanyi echoed this observation, writing, "The organs of the body are more complex and variable than the parts of a machine, and their functions are also less clearly specifiable." He noted that a machine knows only "rules of rightness" and can never accommodate error or "failures"—unlike living bodies that are capable of such radical transformation.[38] Or, as the systems theorist Ludwig von Bertalanffy put it, unlike a living, open system, a "mechanized organism would be incapable of regulation following disturbances."[39]

However, with the rise of cyberneticists came the belief in a wholly new kind of machine technology, a flexible, adaptive one that would mimic the vital improvisation of the organism. Yet the cybernetic machines (examples included self-guided missiles and other servomechanisms) were, it seems, still governed by the logic of automaticity, despite their ability to correct their behavior through negative feedback circuits. Adaptive responses to environmental changes were fully determined by the engineered system. The cybernetic being had specific embedded "goals" and a fixed organization of sensors, processors, and effectors to guide

it to these ends. As a French neurologist put it in 1955, "The rigidity of electrical mechanisms deployed in machines or robots appeared from the start far removed from the variability, the plasticity of those in the nervous system."[40] Was there any possibility of genuine plasticity in the cybernetic machine? Could such a machine exhibit genuine pathologies in the sense that Goldstein and Canguilhem gave the term?[41]

The cybernetician W. Ross Ashby investigated the possibility of such a pathological machine, one that would then be capable of truly novel and unexpected behavior. In a notebook fragment from 1943 we find him reading William James. Ashby was particularly interested in passages from James where the rigid "machine" is compared to the organic and open structure of the nervous system. As we saw, for James this system paradoxically exhibited *both* fixity of structure and an open-ended, adaptive plasticity.[42] The notebooks also show that Ashby was reading Charles Sherrington's revolutionary work on neural integration, published in 1906.[43] Ashby zeroed in on passages that described the dynamic nature of the nervous system. Like James, Sherrington located the great advantage of the nervous system's fragility: by virtue of being so sensitive to disruption, the organism was highly sensitive to changes in its environmental milieu.[44] Inspired by these ideas, Ashby set out to construct a machine that could behave like an open system, behave, that is, like a determinate structure that was nonetheless capable of reorganization.

Ashby admitted in 1941 that artificial machines that could change their organizations were "rare."[45] He was searching for the machine's missing quality, that "incalculable element" (in the words of James) capable of producing novel actions. If machines were defined, as Polanyi had argued, by their calculability, Ashby's project would be impossible. But he eventually hit upon a new way of thinking about machines and plasticity. In essence he realized that for a determinate machine to be radically open to new forms of organization, it had to be capable of becoming, in

a way, a wholly different machine. *Failure*, or breakdown, which was already inevitable in any concrete machine, turned out to be the key idea. For Ashby "a break is a change in organization."[46] A cybernetic machine, one that continually tried to find equilibrium, could at times find itself in a condition where maintaining homeostasis became impossible. However, if that homeostatic being was constructed so that it would, in these moments, break down in some fashion when equilibrium was not being achieved, then (theoretically at least) this machine could be said to acquire a new organization, and therefore new possible paths to equilibrium. As Ashby wrote, "After a break, the organization is changed, and therefore so are the equilibria. This gives the machine fresh chances of moving to a new equilibrium, or, if not, of breaking again."[47] Using the language of neurophysiology in this period, we could say that the internal failure of the machine was a form of shock to the system. Ashby pointed the way, then, to a new form of *cybernetic plasticity* that flowed from the basic weakness of any machine, namely the inevitability of failure.

Ashby's rethinking of the human brain followed from these ideas on the physical forms of adaptive behavior. The brain was not only dynamic and flexible in its organization; it was also a special kind of "machine" made up of individual elements (neurons) that were in essence unreliable, in that they would stop working at certain thresholds until they could recover and begin again to reconnect with other neurons. Due to their latency period, individual neurons were, so to speak, constantly appearing and *disappearing* from the system as a whole. The challenge for cybernetics was to model this form of productive failure within artificially constructed intelligent machines.[48]

The link between brain science and cybernetics offers a new way of contextualizing the origins of AI. Turing opened the field, conceptually at least, when he published "Computing Machinery and Intelligence" in 1950, introducing the famous "imitation game" to the public. Turing's project, as it is usually understood,

was to think about how a digital computer could successfully simulate human intelligence by modeling thought in algorithmic processes. However, in 1946, in the midst of developing Britain's first stored program digital computer, the Automatic Computing Engine, Turing wrote to Ashby. In that letter Turing admitted that he was "more interested in the possibility of producing models of the action of the brain" than in any practical applications of the new digital computers inspired by his work.[49] His interest in models of the nervous system and brain in fact indicate a turn from the strict notion of machine automaticity introduced in his formal mathematical work on computability toward a new interest in the dynamic organization and reorganization of organic systems.

In 1936 Turing had envisioned the universal computer as, in theory, a perfectly automated, stand-alone machine. However, by 1948, after much experience with real computing during and after World War II, he was thinking more and more about the relationship between brains, human intelligence, and these new computers. Turing noted that by definition any machine can be "completely described" by listing all its possible configurations: "According to our conventions the 'machine' is completely described by the relation between its possible configurations at consecutive moments. It is an abstraction that cannot change in time." Yet he imagined that if something *interfered* with this machine, the machine would in essence be modified; it would become a new machine. Turing then suggestively noted that human learners might well be considered "machines" that have been constantly modified by massive interference, namely through teaching and communication, and sensory forms of stimuli.[50]

Turing remarked, "It would be quite unfair to expect a machine straight from the factory to compete on equal terms with a university graduate. The graduate has had contact with human beings for twenty years or more. This contact has been modifying his

behaviour throughout that period. His teachers have been intentionally trying to modify it." The "routines" of thought and action in a student have been "superimposed on the original pattern of his brain." Now it was not the case that the human became an automaton in this process of "programming." Rather the open machine that is subject to such modification by interference is the one that becomes capable of real *creativity*. With new routines the human-machine can, Turing said, "try out new combinations of these routines, to make slight variations on them, and to apply them in new ways."[51] And crucial to genuine intelligence was the freedom from automatic routine: "If a machine is expected to be infallible, it cannot also be intelligent. The machine must be allowed to have contact with human beings in order that it may adapt itself to their standards." Intelligence, Turing wrote, "consists in a departure from the completely disciplined behavior involved in computation."[52] Therefore the intelligence of a computer (or a human) was not measured by its computational prowess but by its radical openness to interference—from outside, most obviously, but also, to a certain extent, from within.

Turing's hypothesis was that the infant human brain should be considered an *unorganized* machine that acquires organization through suitable "interference training": "All of this suggests that the cortex of the infant is an unorganized machine, which can be organized by suitable interference training. . . . A human cortex would be virtually useless if no attempt was made to organize it. Thus if a wolf by a mutation acquired a human cortex there is little reason to believe that he would have any selective advantage."[53] An isolated, solitary human brain would make no progress because it needs a social milieu of other human beings in order to learn and create. Turing's project for AI was not, as we usually think, a project to dissect the operations of cognition and translate them into programmable routines. Instead the goal was to model a mostly open, pliant brain that *transforms itself* as nature and culture impress themselves upon it. "Our hope," he

wrote, "is that there is so little mechanism in the child-brain that something like it can be easily programmed."[54] Human intelligence, like computer intelligence, is in its essence already *artificial*, imprinted, that is, from outside onto the plastic architecture of the brain/computer.

Turing's American counterpart, John von Neumann, was equally fascinated by the robust and flexible behavior of organic nervous systems and brains. Appointed to oversee the development of an electronic computer at the Institute for Advanced Study at Princeton from 1945 to 1951, von Neumann looked to neurophysiology for inspiration. That the brain and the nervous system exhibited an amazing robustness was, he observed, in stark contrast to the immense fragility of the new high-speed computers then being constructed out of undependable mechanical relays, telephone switches, or vacuum tubes. In his early reflections on the computer, von Neumann was careful to draw attention to the critical *differences* between digital machines and the nervous system.[55] Yet he himself was drawn to the nervous system as a way of thinking of the computer as something more than a sequential calculating device. One of the most important marks of the natural communication and control system was, he saw, its inherent flexibility and hence dependability: "It is never very simple to locate anything in the brain, because the brain has an enormous ability to re-organize. Even when you have localized a function in a particular part of it, if you remove that part, you may discover that the brain has reorganized itself, reassigned its responsibilities, and the function is again being performed."[56]

In 1948 von Neumann gave his celebrated lecture "The General and Logical Theory of Automata" at the Hixon Symposium held at Cal Tech. This interdisciplinary gathering on the topic of the brain is one of the key founding moments of cybernetics. Some of the most influential researchers in psychology were in attendance there, including holistic thinkers such as Köhler and Lashley. In fact von Neumann seemed sensitive to this tradition

in his own presentation. He was interested in developing artificial computing machines that mimicked the robustness of natural, living systems, and he pointed to just the kind of nervous system behavior studied by the neurophysiologists. As he explained, the organism was always challenged by unpredictable but inevitable *errors*. How did the nervous system in particular maintain itself despite the pressure of pathological circumstances? In the discussion of von Neumann's paper it was noted that we must "not only be able to account for the normal operation of the nervous system but also for its relative stability under all kinds of abnormal situations."[57]

Error was linked, in von Neumann's work, to plasticity. If a complex computer was going to be flexible and stable across error and failure it was important that "error be viewed . . . not as an extraneous and misdirected or misdirecting accident, but as an essential part of the process under consideration."[58] Somehow the computer would need to have the ability to "automatically" sense where errors were occurring and simultaneously reorganize the system as a whole to accommodate those errors. It seems clear that von Neumann was alluding directly to key concepts taken from theoretical biology, what Lashley called "equipotentiality" and Bertalanffy understood by "equifinality." Von Neumann looked to how natural organisms could *reorganize* themselves when they faced what he labeled "emergency" conditions.[59]

FROM ITS ORIGIN the digital computer was a machine that was always awaiting its determination. To put it another way, the computer needed to contain what the postwar philosopher of technology Gilbert Simondon called "a margin of indetermination." For Simondon the perfect machine could never be perfectly automatic. If it was going to be open to information and therefore capable of genuine communication, then the machine, like the organism, must be in some measure plastic: "The true perfection of machines . . . does not correspond to an increase in

automation, but on the contrary to the fact that the functioning of a machine harbors a certain margin of indetermination." If, as Descartes understood centuries ago, the human being is an automaton of sorts, determined by its education, training, language, and culture, it is nonetheless a special kind of automaton. As Simondon put it, "The human being is a rather dangerous automaton, who is always risking invention and giving itself new structures."[60] The challenge of creating any intelligent automata is the challenge of modeling indetermination, disruption, failure, and error. In an age of automation that figured cognition itself as an automatic procedure, it was doubly important to show that the automaticity of the thinking being was predicated on a more foundational *plasticity*. Lewis Mumford explained this in 1964: "Let me challenge the notion that automation is in any sense a final good. If the human organism had developed on that principle, the reflexes and the autonomic nervous system would have absorbed all the functions of the brain, and man would have been left without a thought in his head. Organic systems [have] . . . the margin of choice, the freedom to commit and correct errors, to explore unfrequented paths, to incorporate unpredictable accidents . . . to anticipate the unexpected, to plan the impossible."[61]

In our own age of automation, where the automaticity incarnated by digital technologies structures the conceptual foundations of cognitive science, it is once again time to rearticulate human nature in terms of what is not automatic. Our digital brains—brains modeled on and simulated by computers and increasingly *formed* by repeated interactions with our digital prostheses—will reveal their genuine plasticity only when they rediscover the power of interrupting their own automaticity.

Endnotes

1 The opening sections of this essay draw on my "Penser l'automaticité sur le seuil du numérique," in *Digital Studies: Organologie des savoirs et technologies de la connaissance*, ed. Bernard Stiegler (Paris: Editions FYP, 2014).

2 Morana Alač, *Handling Digital Brains: A Laboratory Study of Multimodal Semiotic Interaction in the Age of Computers* (Cambridge, MA: MIT Press, 2011).

3 Thomas P. Trappenberg, *Fundamentals of Computational Neuroscience* (Oxford: Oxford University Press, 2010).

4 On the intertwined history of cognitive science and artificial intelligence, see Paul Edwards, *The Closed World: Computers and the Politics of Discourse in Cold War America* (Cambridge, MA: MIT Press, 1997); Jean-Pierre Dupuy, *On the Origins of Cognitive Science: The Mechanization of Mind*, trans. M. B. DeBevoise (Cambridge, MA: MIT Press, 2009).

5 Human Brain Project, "Overview," https://www.humanbrainproject.eu/discover/the-project/overview (accessed February 12, 2015).

6 U.S. Department of Health and Human Services, National Institutes of Health, "Brain Research through Advancing Innovative Neurotechnologies (BRAIN)," braininitiative.nih.gov (accessed February 12, 2015).

7 Catherine Malabou, *The Ontology of the Accident: An Essay on Destructive Plasticity*, trans. Caroline Shread (London: Polity, 2012).

8 Rajagopal Ananthanarayanan, Steven K. Esser, Horst D. Simon, and Dharmendra S. Modha, "The Cat Is Out of the Bag: Cortical Simulations with 10^9 Neurons, 10^{13} Synapses," *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*, article 63, http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6375547 (accessed June 2, 2015).

9 Paul Merolla et al., "A Digital Neurosynaptic Core Using Embedded Crossbar Memory with 45pJ per Spike in 45nm," *IEEE Custom Integrated Circuits Conference*, September 2011. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6055294 (accessed June 2, 2015).

10 Mrigank Sharad et al., "Proposal for Neuromorphic Hardware Using Spin Devices," Cornell University Library, July 18, 2012, http://arxiv.org/abs/1206.3227v4 (accessed June 2, 2015).

11 Dharmendra S. Modha et al., "Cognitive Computing," *Communications of the ACM* 54 (2011): 65.

12 B. G. Farley and W. A. Clark, "Simulation of Self-Organizing Systems by Digital Computer," *Transaction of the IRE* (1954): 7.

13 René Descartes, *Treatise on Man*, in *The Philosophical Writings of Descartes*, trans. John Cottingham, Robert Stoothoff, and Dugald Murdoch, 3 vols. (Cambridge: Cambridge University Press, 1985), 1: 104.

[14] René Descartes, *Discourse on Method*, in *Philosophical Writings*, 1: 119.

[15] David Bates, "Cartesian Robotics," *Representations* 124 (Fall 2013): 43–68.

[16] C. S. Peirce, *Collected Papers of Charles Sanders Peirce*, 8 vols. (Cambridge, MA: Harvard University Press, 1931–58), 2: para. 59.

[17] C. S. Peirce, letter to A. Marquand, December 30, 1886, in C. Kloesel et al., eds., *Writings of Charles S. Peirce: A Chronological Edition* (Bloomington: Indiana University Press, 1993), 5: 421–22. An image of the original letter with the circuit design diagram is on 423.

[18] William James, *Principles of Psychology*, 2 vols. (New York: Henry Holt, 1890), 1: 105.

[19] Ibid., 1: 139–40, 143.

[20] C. S. Peirce, *Contributions to the Nation*, 3 vols. (Lubbock: Texas Tech University Press, 1975–87), 1: 144.

[21] Constantin von Monakow, *Die Lokalisation im Grosshirn und der Abbau der Funktion durch kortikale Herde* (Wiesbaden: J. F. Bergmann, 1914), 30.

[22] Constantin von Monakow and R. Mourgue, *Introduction biologique à l'étude de la neurologie et disintegration de la function* (Paris: Félix Alcan, 1928), 29.

[23] The following discussion of neurophysiology takes up the argument made in my "Unity, Plasticity, Catastrophe: Order and Pathology in the Cybernetic Era," in *Catastrophes: History and Theory of an Operative Category*, ed. Andreas Killen and Nitzen Lebovic (Berlin: De Gruyter, 2014), 32–54.

[24] Karl Lashley, *Brain Mechanisms and Intelligence: A Quantitative Study of Injuries to the Brain* (Chicago: University of Chicago Press, 1929), 176.

[25] Karl Lashley, "Persistent Problems in the Evolution of Mind," *Quarterly Review of Biology* 24 (1949): 33, 31. The Soviet psychologist and neurologist Alexander Luria developed a similar theorization of neural plasticity. In his studies of patients who had suffered various brain injuries, Luria observed the loss of certain cognitive abilities, but he also observed repeatedly that the brain had the startling ability to reorganize itself in order to compensate for the loss of functions after stroke or accident. Luria noted that numerous studies showed "the high degree of plasticity shown by damaged functional systems, due to dynamic reorganization and adaptation to new circumstances and not to regeneration and restoration of their morphological integrity." A. R. Luria, *Restoration of Function after Brain Injury*, trans. Basil Aigh (New York: Macmillan, 1963), 33. See as well Laura Salisbury and Hannah Proctor's essay in this volume.

[26] Kurt Goldstein, *The Organism: A Holistic Approach to Biology Derived from Pathological Data in Man* (1934; New York: Zone, 1995), 48–49. An excellent account of Goldstein's work and the German contexts of theoretical biology can be found in Anne Harrington, *Reenchanted Science: Holism in German Culture from Wilhelm II to Hitler* (Princeton, NJ: Princeton University Press, 1996).

[27] Goldstein, *The Organism*, 52, 35.

[28] Ibid., 227, 388.

[29] Wolfgang Köhler, *Intelligenzprüfungen an Menschenaffen* (Berlin: Springer, 1921).

[30] Wolfgang Köhler, "Some Gestalt Problems," in *A Sourcebook of Gestalt Psychology*, ed. Willis D. Ellis (London: Kegan Paul, 1938), 55, 66.

[31] Karl Duncker, *Zur Psychologie des produktiven Denkens* (Berlin: Springer, 1935); Max Wertheimer, *Productive Thinking* (Ann Arbor: University of Michigan Press, 1945).

[32] Kurt Koffka, *The Growth of the Mind: An Introduction to Child Psychology* (New York: Harcourt Brace, 1925), 125.

[33] Edouard Claparède, "La genèse de l'hyopthèse: Étude expérimentale," *Archives de Psychologie* 24 (1934): 5.

[34] Vladimir Jankélévitch, *Henri Bergson* (Paris: Presses Universitaires de France, 1989)

[35] Warren McCulloch and Walter Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity," *Bulletin of Mathematical Biophysics* 5 (1943): 115–33; Norbert Wiener, *Cybernetics* (Cambridge, MA: MIT Press, 1948); Warren McCulloch and John Pfeiffer, "Of Digital Computers Called Brains," *Scientific Monthly* 69 (1949): 368–76.

[36] Georges Canguilhem, "Machine et organisme" (1948), in *La connaissance de la vie*, 2nd ed. (Paris: Vrin, 1980), 118.

[37] Georges Canguilhem, *On the Normal and the Pathological*, trans. Carolyn R. Fawcett (New York: Zone, 1991).

[38] Michael Polanyi, *Personal Knowledge: Towards a Post-Critical Philosophy* (Chicago: University of Chicago Press, 1958), 359, 328.

[39] Ludwig von Bertalanffy, *Das Biologische Weltbild* (Berne: A. Franke, 1949), 29.

[40] Théophile Alajouanine, *L'homme n'est-il qu'un robot? Considérations sur l'importance qu'a l'automaticité dans les functions nerveuses. Discours inaugural prononcé à la séance du Congrès des psychiatres et neurologistes de langue française, à Nice, le 5 septembre 1955* (Cahors: A. Coueslant, 1955), 11.

[41] We can note that Bertalanffy had criticized cybernetics precisely for its misguided use of closed systems to model the fundamentally open structures of natural organisms. Ludwig von Bertalanffy, *General System Theory: Foundations, Development, Applications* (New York: George Braziller, 1968).

[42] W. Ross Ashby, *Journal*, 1523–24, W. Ross Ashby Digital Archive, rossashby.info/index.html (accessed July 24, 2013).

[43] Charles S. Sherrington, *The Integrative Action of the Nervous System* (New Haven, CT: Yale University Press, 1906).

[44] Ashby, *Journal*, 1906–9.

[45] Ibid., 1054.

[46] W. Ross Ashby, "The Nervous System as Physical Machine: With Special Reference to the Origin of Adaptive Behaviour," *Mind* 56 (1947): 50. On Ashby's theoretical work and engineering projects in homeostasis, see Andrew Pickering, *The Cybernetic Brain: Sketches of Another Future* (Chicago: University of Chicago Press, 2009), especially chapter 4

[47] Ashby, "The Nervous System as Physical Machine," 55.

[48] Ibid., 57–58.

[49] Alan Turing, letter to W. Ross Ashby, ca. November 19, 1946, quoted in Jack Copeland, introduction to Alan Turing, "Lecture on the Automatic Computing Engine (1947)," in Alan Turing, *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life plus the Secrets of Enigma*, ed. B. Jack Copeland (Oxford: Oxford University Press, 2004), 374.

[50] Alan Turing, "Intelligent Machinery" (1948), in Turing, *The Essential Turing*, 419.

[51] Ibid., 421.

[52] Alan Turing, "Computing Machinery and Intelligence," *Mind* 59 (1950): 459.

[53] Turing, "Intelligent Machinery," 424.

[54] Turing, "Computing Machinery," 456.

[55] John von Neumann, *The Computer and the Brain* (New Haven, CT: Yale University Press, 1958).

[56] John von Neumann, *Theory of Self-Reproducing Automata*, ed. Arthur W. Burks (Urbana: University of Illinois Press, 1966), 49.

[57] Ibid., 323.

[58] Ibid., 329. I have discussed von Neumann's theory of error in "Creating Insight: Gestalt Theory and the Early Computer," in *Genesis Redux: Essays in the History and Philosophy of Artificial Life*, ed. Jessica Riskin (Chicago: University of Chicago Press, 2006), 237–59. See as well Giora Hon, "Living Extremely Flat: The Life of an Automaton. John von Neuman's Conception of Error of (In)animate Systems," in *Going Amiss in Experimental Research*, ed. Giora Hon, Jutta Schickore, and Friedrich Steinle, Boston Studies in the Philosophy of Science, no. 267 (N.p.: Springer, 2009), 55–71.

[59] Von Neumann, *Theory of Self-Reproducing Automata*, 71, 73.

[60] Gilbert Simondon, *Du mode d'existence des objets techniques*, rev. ed. (Paris: Aubier, 1989), 11–12.

[61] Lewis Mumford, "The Automation of Knowledge: Are We Becoming Robots?," *Audio-Visual Communication Review* 12 (1964): 275.